JTLU

An integrated microsimulation approach to land-use and mobility modeling

Yi Zhu
Shanghai University of Finance and Economics
zhu.yi@mail.shufe.edu.cn

Mi Diao
National University of Singapore
rstdm@nus.edu.sg

Joseph Ferreira, Jr.
Massachusetts Institute of Technology
jf@mit.edu

P. Christopher Zegras
Massachusetts Institute of Technology
czegras@mit.edu

Abstract: This paper presents an overview of the design and status of a new type of land-use simulation module integrated into SimMobility, an agent-based microsimulation platform. The module, SimMobility Long-Term (LT), is designed to simulate how the interrelations between the transportation and land-use systems manifest themselves in the housing and commercial real estate markets, household and firm location choices, school and workplace choices, and vehicle ownership choices. At the heart of the LT simulator is a housing market module simulating daily dynamics in the residential housing market that (a) "awakens" households that begin searching for new housing; (b) accounts for eligibility, affordability, and screening constraints; (c) constructs plausible choice sets; (d) simulates daily housing market bidding; and (e) represents developer behavior regarding when, where, what type, and how much built space to construct, taking into account market cycle and uncertainty. The LT simulator and SimMobility's activity-based, mid-term (MT) simulator are integrated in that the agents in the LT module (e.g., individuals in households) are the same agents simulated in the MT module (e.g., activity participation and trip-making), and agents' simulated behaviors in the MT module provide measures of (utility-based) accessibility that figure into relevant decisions in the LT simulator.

This paper describes the SimMobility model and the LT framework, presents estimation results for two component models of the housing supply side, and demonstrates the use of the simulator by comparing housing market outcomes with and without the introduction of one year's worth of supply of new public housing. Overall, the LT simulator represents an effort to advance urban system modelling by explicitly simulating the dynamic interactions of disaggregated agents in real estate markets and encapsulating the information of agents' daily activity participation in their long-term mobility-relevant choices.

1 Introduction

Among the relevant policy areas for effectively managing metropolitan growth, coordination of metropolitan land use and transportation planning has long been recognized as important. The spatial distribution of people, places, and firms determines the locations of potential trip origins and destinations and influences the relative attractiveness of different travel modes. Transportation infrastructures, services and demands, in turn, influence the relative desirability of different places, influencing the physical growth of the city. Predicting the impacts of these interactions, such as zoning policy effects on traffic congestion or new rail investment impacts on land development, is important to researchers and practitioners alike. Since the early days of the Lowry model, operational models linking, if not integrating, these two urban sub-systems have been developed and applied as an aid to relevant decision-making. In practice, however, the use of such models tends to remain the exception, in some sense for the same reasons identified in Lee's (1973) seminal "Requiem for Large-Scale Models."

Today, however, a convergence of factors, partly enabled by advances in information and communication technologies, make integrated modeling highly relevant to decision-making and, potentially, more feasible in practice. The dynamics of the systems of interest are seemingly ever-increasing in complexity: for example, the heterogeneous spatiotemporal activity patterns of individuals, households and firms are driving new dynamics between, and services for, passenger and freight demand and supply, with important possible implications for location choices (e.g., homes, workplaces). Investments, policies and pricing can, in theory, be more nuanced and sophisticated, targeting, for example, times of day and specific user groups, and adopting flexibilities (e.g., in road-space allocation) previously difficult to achieve. Finally, ongoing improvements in computing power, data availability and behavioral modeling make the design and implementation of behaviorally rigorous, agent-based microsimulation tools increasingly feasible.

Nearly 15 years ago, Wegener (2004) identified microsimulation of integrated activity-based land-use/transport models as the "most promising technique" for representing the complex spatial behavior of the modern metropolis. On the transportation modeling side, activity-based modeling and microsimulation (Castiglione, Bradley, & Gliebe, 2015; Ben-Akiva, Bowman, and Gopinath, 1996) are becoming increasingly common, providing advantages in terms of behavioral rigor and policy sensitivity compared with the conventional four-step travel model approach (Bhat & Singh, 2000; Buliung & Kanaroglou, 2007). On the land-use side, recent years have seen a turn to disaggregated and dynamic models as opposed to aggregate and steady-state equilibrium approaches (Waddell et al., 2003; Miller, Hunt, Abraham, & Salvini, 2004). Nonetheless, only modest progress has been made in consistently integrating disaggregate, microsimulation- and activity-based models of transport and land use.

This paper describes recent efforts to develop a functionally integrated microsimulation model: SimMobility. SimMobility integrates the land use, activity and transport systems across three time frames, roughly consistent with decision-making: short-term (ST), e.g., lane changing behavior; medium-term (MT), e.g., activity and mode choices; and long-term (LT), e.g., vehicle ownership and housing choice. This paper focuses on the LT model system and its integration with MT. The remainder of the paper is organized as follows. Section 2 describes the overall theoretical foundation underlying SimMobility's land use and transportation integration. Section 3 describes the LT model framework, including the deployment environment and simulation component. Section 4 presents the LT models, describing several in detail. Section 5 presents empirical estimation results from part of the real estate supply model. Section 6 presents a simulation example for a Singapore case that includes, or excludes, one year's worth of new public housing. Section 7 covers conclusions and future work.

2 Integrating land use and transport via utility-based accessibility

The concept of accessibility provides the fundamental *economic* connection between the land-use and mobility subsystems (Martínez, 1995). Accessibility can be measured in many different ways (e.g., Geurs, & van Wee, 2004), yet when conceptualized as this formal economic land-use and transport link, economically consistent accessibility measures are theoretically important for model integration. In this sense, utility-based measures of accessibility, derivable from random utility (i.e., discrete choice) models, emerge as particularly attractive. Not only can they reflect individual preferences (based on the individual's actual choice), but they are directly linked to traditional measures of consumer surplus (e.g., Small & Rosen, 1981). Ben-Akiva and Lerman (1979) first explicitly tie the disaggregate discrete choice modeling framework to the accessibility concept, defining the expected maximum utility (i.e., the denominator of the logit model, representing the value of an individual's entire choice set) as a "reasonable" accessibility measure.

More formally, in the discrete choice framework, utility, *U*, consists of two, additive components:

$$U_{jn} = V(z_{jn}, s_n, \beta) + \varepsilon_{jn} \tag{1}$$

where *V* represents a systematic utility function, z_{jn} is a vector of attributes of the alternatives *j* for decision-maker *n*, s_n is a vector of socioeconomic and/or demographic characteristics of the choice-maker, β is a vector of unknown parameters and ε_{jn} represents the unobservable, unknown portion (i.e., the random "disturbance") of utility. In the widely familiar *logit* model of choice, the error term is assumed to take on an identical, independent distribution (IID), that is also Gumbel-distributed with a scale parameter, *μ*. The resulting, basic probability framework becomes:

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j=1}^{j} e^{\mu V_{jn}}} \tag{2}$$

The denominator in equation 2 represents the expected maximum utility (or "logsum," from its form) from the set of relevant alternatives:

$$E(\max_{i \in C_n} U_{in}) = \frac{1}{\mu} \ln \sum_{i \in C_n} e^{\mu V_{in}} \tag{3}$$

where V_{in} is the systematic component of utility U_{in} for individual *n* choosing one alternative from the choice set C_n. This term serves as a summary measure or index of the utility of the entire choice set and is what Ben-Akiva and Lerman (1979) call a "measure of accessibility." In the case of the more general nested logit model (NL), the logsums "pass up" the choice chain, with the logsums from the lower level choices (e.g., mode choice), being included within the utility component of higher levels (e.g., destination choice), up to the root, or highest level (e.g., activity pattern choice). The logsum calculated for the root represents the expected value to the individual of the full choice set.

Building on this framework, Ben-Akiva and Bowman (1998) outline the theoretical basis for integrating activity-based models of travel demand with residential location choices and demonstrate estimation results in a nested logit model system. In their example, activity-based accessibility (ABA) represents an individual's maximum utility from her available activity schedules (combinations of activities, destinations, modes), given a residential location. The approach: allows for different residential locations

to have different accessibilities for different individuals; reflects individual differences in preferences for activity schedules (including modes and destinations); and, can reflect the possibilities for activity substitution, trip chaining, and other behaviors that might be influenced by variations in location choices (Ben-Akiva & Bowman, 1998). In other words, the ABA provides a potentially rich, behaviorally based, way of representing the interactions between longer-term choices, such as where an individual might choose to live, and that choice's relationship to an individual's preferred activities, when and where she does them, and how she prefers to travel to them. In theory, the approach can be extended to other relevant decisions, such as vehicle ownership, and to decisions by other agents, such as a developer's choice of where to build housing or a firm's location choice.

While theoretically attractive, the integration of activity-based models with land-use/development models, microsimulation-based or otherwise, remains rare in practice (Kii, Nakanishi, Nakamura, & Doi, 2016). For example, ILUMASS, an integrated land-use/transport microsimulation model system, did not ultimately fully integrate its detailed activity-based transport microsimulation model into the land-use model, due to model complexity and computational constraints (Wagner & Wegener, 2007). ILUTE, an integrated urban system microsimulation model includes a travel/activity simulator (Salvini & Miller, 2005), but activity-based accessibility measures do not feed directly into relevant land-use submodels. A Statewide integrated model for Oregon (USA) (SWIM) has an activity-based person-travel model system which links to a spatial activity allocation model, but only using logsum-based accessibility from the mode choice model (Donnelly, Upton, & Knudson, 2018). UrbanSim, a dynamic land-use micro-simulation model (Waddell, 2000) has been linked with several different activity-based microsimulation models. For example, SimTRAVEL (Simulator of Transport, Routes, Activities, Vehicles, Emissions, and Land), an integrated activity-based microsimulation tool, used UrbanSim as its landuse model; the prototype application used time-dependent auto network performance for accessibility (Pendyala, et al., 2012). Nicolai and Nagel (2012) integrate UrbanSim with MATSim, a disaggregate agent-based transport model. That application also did not use an ABA measure, instead they calculate zone-level accessibility using shortest path-based travel costs on the road network. Other researchers have also used MATSim in integrated land-use/transport models. Ziemke, Nagel, and Moeckel (2016), integrated MATSim with a microscopic land-use model, SILO, using MATSim-based rush hour travel times to compute zone-level accessibilities for use in SILO's household relocation modelling. Zhuge, Shao, Gao, Dong, and Zhang (2016) incorporate MATSim into SelfSim, which includes a microsimulation-based residential location choice and real estate price model. They use a type of accessibility-based indicator generated by MATSim, a sum of the utilities of all activities performed by an individual (e.g., Charypar & Nagel, 2005). This measure is based on an agent's predicted travel plan, since MATSim does not calculate something similar to the expected value of the entire choice set.

In summary, while research remains active in the development of integrated activity-based land-use/transportation microsimulation tools, few, if any, examples exist of integration using the full economic information theoretically represented by the logsum of an activity-based travel demand model. SimMobility is a simulation tool based on this key integration principle.

3   SimMobility simulation platform

SimMobility is a multi-agent based, integrated microsimulation platform being developed primarily by the Future Urban Mobility research group under the Singapore-MIT Alliance for Research and Technology (SMART). SimMobility, integrates three main modules, segmented according to the simulation timeframes (see Figure 1):

- Short-term (ST)—The basic time step is a tenth of a second and agent decisions include lane changing, braking, accelerating, and gap acceptance. The ST is essentially a microscopic traffic simulator, extended with models for pedestrians and public transportation.
- Mid-term simulator (MT)—An activity-based model, MT's time step ranges from seconds to minutes and agent choices include pre-day activity plan, route (and re-routing), mode, activity (re)scheduling, and departure time (Lu et al., 2015).
- Long-term simulator (LT)—The basic time step is one day and agent decisions include choices related to housing location, job location, land development, and vehicle ownership.

The SimMobility framework represents a behaviorally consistent, temporally integrated set of key decisions and decision-makers involved in relevant land use and transportation interactions. Modules are stand-alone and have built-in functions facilitating information exchange with other modules. Across the timescales, SimMobility follows an event-driven, activity-based paradigm, simulating supply and demand at each level and the interactions between different levels. For example, LT provides population characteristics and land-use configuration to the MT, which transmits trip-chains to ST. The ST provides performance measures to MT, which provides accessibility measures (logsum-based ABA; i.e., following equation 3) to the LT. The model system is designed for simulating both person and goods movements, although the latter set of models has received less research attention to date. Adnan et al. (2016) provide additional detail.
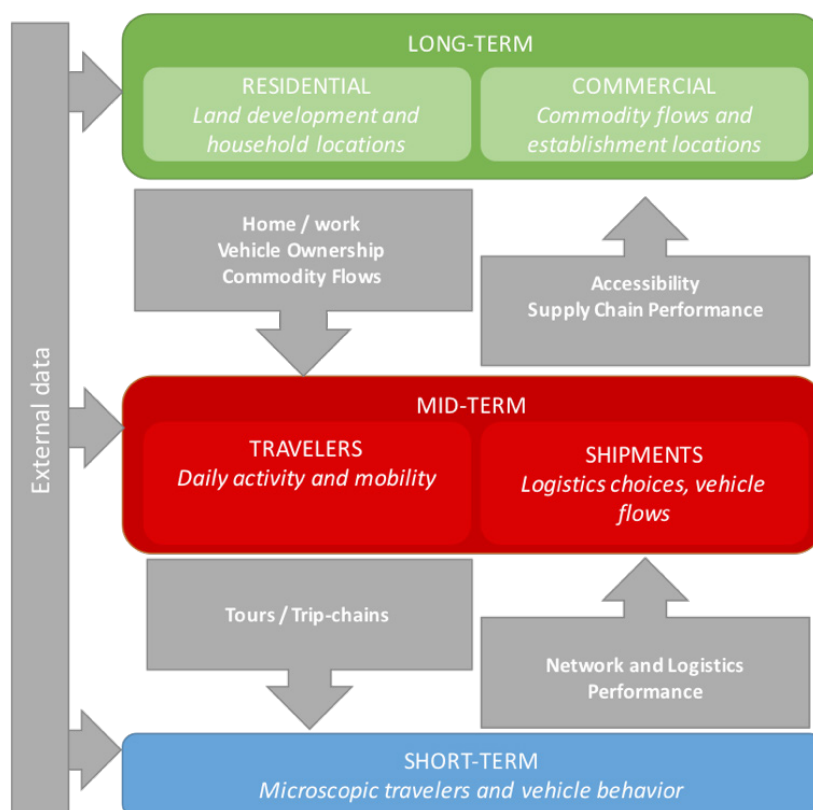

Figure 1: SimMobility modeling framework

Beyond the theoretical attractiveness of its integrated framework, SimMobility's architectural integration offers various advantages. With a single database, for example, demand can be consistently modeled at the disaggregate level, while tracking individualized history with relatively modest burden. Developed in a common programming language (C++), SimMobility avoids some of the challenges of integrating

softwares developed in different languages (e.g., MATSim and UrbanSim; see Ziemke et al., 2016) and capitalizes on a parallel and distributed architecture intended to maximize computational efficiency and increase scalability.

4   Long-term simulation framework

Currently, the operational elements of SimMobility LT primarily involve the residential housing market, as depicted in Figure 2.[1] The simulation begins with a "day-0" estimate of the characteristics and spatial location of the synthesized residential population, residential units by housing type, firms, establishments, and commercial spaces.[2] Each household is initially assigned to a particular housing unit in a specific building. Unoccupied units available for sale constitute the housing market supply and consist of resales and new sales (including units available for advance purchase[3]). Sellers set asking prices a few percentage points above the perceived market price (based on a hedonic price model). Rather than model sales and price adjustments as quarterly or annual events, LT models housing market transactions as a daily bidding process among those buyers and sellers estimated to be actively engaged in searching for housing and negotiating sales[4].

Potential buyers come from the entire pool of households based on explicit probabilistic models for "awakening," determining affordability and eligibility, and "screening" unit availability by housing and neighborhood type. These submodels—combined with a simple supply-side model of property resales and a more sophisticated model of (re)development—reduce the number of buyers and sellers active on any given day. This makes it practical to construct plausible choice sets for all active households and simulate a daily bidding process. Active buyers evaluate each housing unit in their choice set and then bid on the unit maximizing their expected surplus. The following section provides more detail. Once successful buyers move into their new units, they reassess the job and school assignments of their household members, and reconsider vehicle-ownership[5]. The suppliers' development choice model, and the households' bidding, job/school location choice and vehicle ownership choice all include the ABA measures from MT (equation 3).

---

[1]We model rentals as well as owner-occupied housing units and households can transition between renting and owning. This paper only discusses the housing market sales model. The rental market model is not yet fully implemented.

[2]The synthetic population of firms, establishments, and commercial spaces is necessary to locate jobs, schools, and non-work destinations. The market model for commercial real estate, however, is not yet fully implemented.

[3]In the case of Singapore, new housing units may be offered for sale by private developers or the government up to four years before they are completed and ready for occupancy.

[4]To our knowledge, SimMobility is the only microsimulation land-use and transportation model that has implemented a daily bid-auction model for housing in a large metropolitan area using econometrically estimated models of housing price and accessibility.

[5]At the moment, the job, school, and vehicle choices are handled sequentially and are not bundled with the housing choice. They can be triggered by events other than moving (such as losing one's job, or responding to changes in interest rates or product marketing) but they are not evaluated simultaneously when bidding for housing units.
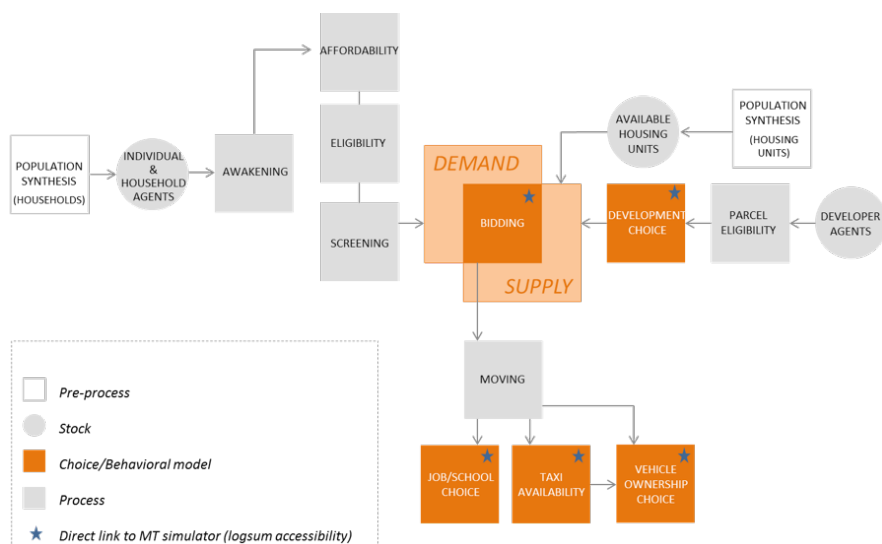
Figure 2: Flow chart of daily bidding activity in the residential real estate market simulator

4.1 Initializing the agent population

SimMobility has four agent types representing "real world" households, individuals, firms and establishments. These agents make mobility- and location-related decisions, in light of their current states and conditions. Agents' attributes intend to sufficiently distinguish them and capture relevant heterogeneities, subject to data availability. Relationships among different agent types (such as individual-household linkages) are retained in SimMobility, which enables the simulation to capture inter-agent relationships, such as intra-household influences on decisions like home location choice and daily activity scheduling.

A two-stage data synthesis method is employed to build a synthetic population containing all individuals, households and establishments in the simulated urban system with their important locations (residences and workplaces) assigned to units/buildings (drawing from available spatial information about building characteristics and spatial variation in population characteristics). The detailed socio-economic and demographic attributes of this synthetic population are generated using a customized iterative proportional fitting (IPF) method that minimizes the discrepancy between the attribute distributions and correlations in the synthetic population and those observable in marginal distributions recorded in official statistics (Zhu & Ferreira, 2014, 2015). This high spatial resolution enables the simulation to run at a quite disaggregated spatial level from the start. Table 1 lists the most important agent attributes obtained from official statistics, and included as marginal constraints. The population synthesis process was able to generate 5,077 different household types and 1,398 individual types for the 1+ million households assigned to 330 residential subzones during the first stage of the population synthesis process.

At the end of each simulation year, an additional agent synthesis process adjusts the agent population to accommodate exogenous changes in migration, economic conditions, etc. This process must avoid undoing the cumulative effects incurred from the endogenous microsimulation steps. For example, the stipulated annual percentage population growth is not distributed uniformly across residential communities, but located using the housing market model (Figure 2) during the course of each simulated year. Likewise, household agents are designed to retain limited memory (about years living in the same house, the geographic location of one's job, etc.), and inter-agent relationships may be constrained for a few simulation time to add stability and a realistic built-in inertia to the simulation.

Table 1: Attributes of agents

Individual	Household	Establishment
Ethnicity*	Size*	Business type*
Employment status	# Children*	Employment size*
Gender*	# Workers*	Revenue
Education	Income *	Postcode address*
Occupation	Time in current housing	
Income	Housing unit type*	
Age*	Vehicle ownership	
Citizenship status*		
Household head		

* Indicates variables from official statistics that are used as marginal controls when constructing the synthetic population

4.2   Spatial details

The LT simulator explicitly represents buildings, parcels, traffic analysis zones (TAZs) and planning zones, as well as their geographical relationships. The hierarchical representation of spatial elements is required by the processing of spatial data from multiple sources, but is also dictated by the different decision making processes simulated. Households make choices of home location at the level of residential unit; the "choice set" they evaluate is probabilistically constructed from all for-sale housing units based on the likelihood that a particular unit's type and location is a good match. The logsum measures of accessibility differ by socio-economic factors and TAZs, but the TAZ values come from simulating each individual's building-to-building trips. The second stage of the synthetic population generation takes the population (and firm establishments) that were allocated to the 330 sub-zones in the first stage and assigns them to particular housing units in residential buildings (and firm establishments in commercial space) based on the estimated amount and type of floor space available in each building (Zhu & Ferreira, 2015).

Dwelling units, buildings and parcels constitute LT's main spatial object system. Residential buildings contain dwelling units, which have attributes including unit type (1-room public housing flat, apartment, etc.), floor area, tenure, and story. In commercial buildings, floor area is differentiated depending upon whether it is below ground, at or a few stories above ground, or higher up. Buildings are located on land parcels, which contain information on land-use type and floor area ratio allowed by the land-use plan. Importantly, LT records the starting and ending time points of spatial objects' status. For example, residential units have physical status (under construction, finished, demolished, etc.) and sale status (new, listed, transacted, etc.). This allows the simulator to track spatial objects' temporal evolution under a variety of scenarios. The LT simulator also manages amenities (e.g., bus stops, subway stations, malls, schools and highway entrances) that help characterize the built environment at different levels of spatial detail.

Table 2: Attributes and representation of space

Spatial	Attributes	Representation	Major models
Housing units	Size, level, type etc.	Point	Housing bidding and real estate price estimation
Postcodes	A group of urban amenity indicators (e.g., distance to the nearest bus stop) are calculated at postcode level	Point	
Buildings	Age, floor, footprint area, etc.	Polygons	Locations of establishments, jobs, etc.
Land parcels	Land tenure type. Area, land-use type, etc.	Polygons	Real estate development
Traffic analysis zones	Name, code, area, planning area, etc.	Polygon	Population synthesis, logsum measures, results for comparison and validation
Planning area	Name, region, etc.	Polygon	Population synthesis

4.3   Temporal representation

Although time is a continuous variable in the real world, the simulation requires some form of discretization to reduce complexity and computational cost. The LT simulator uses a hybrid approach that relies on both time-steps and events. The LT simulator runs on a daily time period, with real estate transactions, vehicle purchases, job changes, etc. happening at "midnight" of each simulated day. This approach reduces the scale of the bid-auction submodel to a computationally tractable one that retains a structure that can be calibrated from survey data. In Singapore, for example, we model more than one million households (excluding certain construction workers, work permit holders, and foreign households). However, fewer than 100,000 units are sold each year and we estimate that less than one percent of the households are actively searching for new housing on any particular day. Modeling daily transactions also facilitates the direct simulation of certain market disruptions such as changing interest rates, new government housing programs, and car-lite neighborhood initiatives.

The current version of the LT simulator also supports variable time increments that allow processes and objects to be updated at their natural temporal frequency. For example, residential moves resulting from housing purchases (or rentals) may be delayed by up to a few months (or even years for advance purchase of new units) to reflect typical delays in transaction commitments and physical moves. Compared with the housing market module, the time step of land (re)development is relatively longer (e.g., monthly or quarterly). At each time step, the real-option element (described further below) of the model helps to adjust developers' decisions on the timing of different stages of developments, such as construction and selling.

Many other temporal attributes exist to process and update agents and their decisions on a flexible but trackable schedule. For example, date of birth reveals individuals' ages. Households have attributes, such as "unit pending" and "pending occupancy date," to handle the situation of advanced purchase of new dwelling units. These types of temporal leads and lags are usually associated with future events in the simulation, and can be readily handled by an Event Publisher-Subscriber mechanism. For example, households not currently "awakened" may subscribe to a service that informs them (at midnight) of significant interest rate changes or public housing announcements and subsidies for which they are eligible. The day after any such announcement, their chances of awakening can be increased so that they

can more realistically respond via their active participation in the housing market. Similar "publish and subscribe" services can notify other agents, such as sellers and developers agents, about particular types of events which may then result in behavior adjustment on subsequent days.

5   Real estate market models and accessibility

Table 3 describes LT's key sub-models (as of April 2018) along with the data used to estimate their parameters, the subset of agents to which they apply, the relevant spatial and temporal scales, and their use (if any) of accessibility measures obtained from the mid-term simulator. Figure 3 depicts the housing market simulation steps that implement and integrate these sub-models. A major emphasis of the LT simulator is the capability to simulate the dynamic demand-supply interactions of residential and commercial real-estate markets in a manner that explicitly considers the accessibility of residential and commercial locations. For example, the supply of new real estate housing units comes from a model of developer decisions about land (re)development. This section details the modelling approach of key parts of the real estate market model system and describes the current approach to incorporating ABA measures in relevant LT sub-models.

Table 3: Summary of sub-models in the SimMobility LT simulator for Singapore

| Model	Description	Active Agent	Choice Set	Accessibility Measure	Temporal Unit	Spatial Unit
Household awakening	Probabilistic model of moving & tenure transition rates. Estimated from census & recent mover survey	All households	Whether to consider moving and changing tenure		Daily	
Household eligibility/affordability	Rule-based model to estimate household's housing budget and subsidy eligibility	All awakened households	Residential unit types	-	Daily	-
Housing screening	Multinomial Logit (MNL) model of likelihood of residing in each of 7 housing unit types by 35 planning districts. Estimated from census, neighborhood amenities, household travel survey.	All awakened households	35 Planning areas by 7 housing unit types	Person-specific activity-based accessibility (ABA) aggregated to TAZ level	Daily	Planning areas
Housing bidding	Willingness-to-pay model and bidding strategy to identify unit that maximizes expected consumer surplus. Estimated from real estate transactions, travel survey, neighborhood amenities, census.	All awakened households	Available residential units on the market.	Person-specific ABA aggregated to TAZ level	Daily	Dwelling unit
Real estate price estimation	Hedonic model estimating the market price (and asking price) of real estate properties by types. Estimated from census data, neighborhood amenities, real estate transactions.	Developers / Sellers	Land parcels eligible for new and re-development	Person-specific ABA aggregated to TAZ level	Monthly	Parcel
Development template choice	Probabilistic model selecting development template based on expected profitability. Estimated from real estate transactions, land sales, macroeconomic indicators, construction costs.	Developers	Appropriate development templates	-	Monthly	Parcel
Taxi access	Binary logit model estimating the possibility a household includes a taxi driver. Estimated from government statistics and household travel survey.	All households with working household members	Yes or no	-	Yearly & after moving	-
Job assignment	MNL model assigning jobs to workers by matching industry sectors. Estimated from household travel survey.	All types of workers who change jobs or residence.	Sample of workplace postcodes including current workplaces.	Person-specific ABA: fixed residential location, varying workplaces.	Monthly	Building/postcode
School assignment	MNL model assigning students to schools by types of schools (primary, secondary, university etc.).	Full-time students; Kindergarten & Primary students who change home location	Schools/school buildings	Vehicle travel time/Public transit time among ODs	Once a year (start of fall semester)	Building / postcode
Vehicle ownership	MNL model of household vehicle ownership (0,1,2+ cars; motorcycles; off-peak cars) conditional on whether household has primary taxi driver. Estimated from household travel survey	Households who change home location	6 options for 0, 1, 2 and 2+ cars (plus motorcycle and off-peak options)	Household specific ABA: residential unit and current workplace(s), varying vehicle ownership options	Yearly & after moving	-
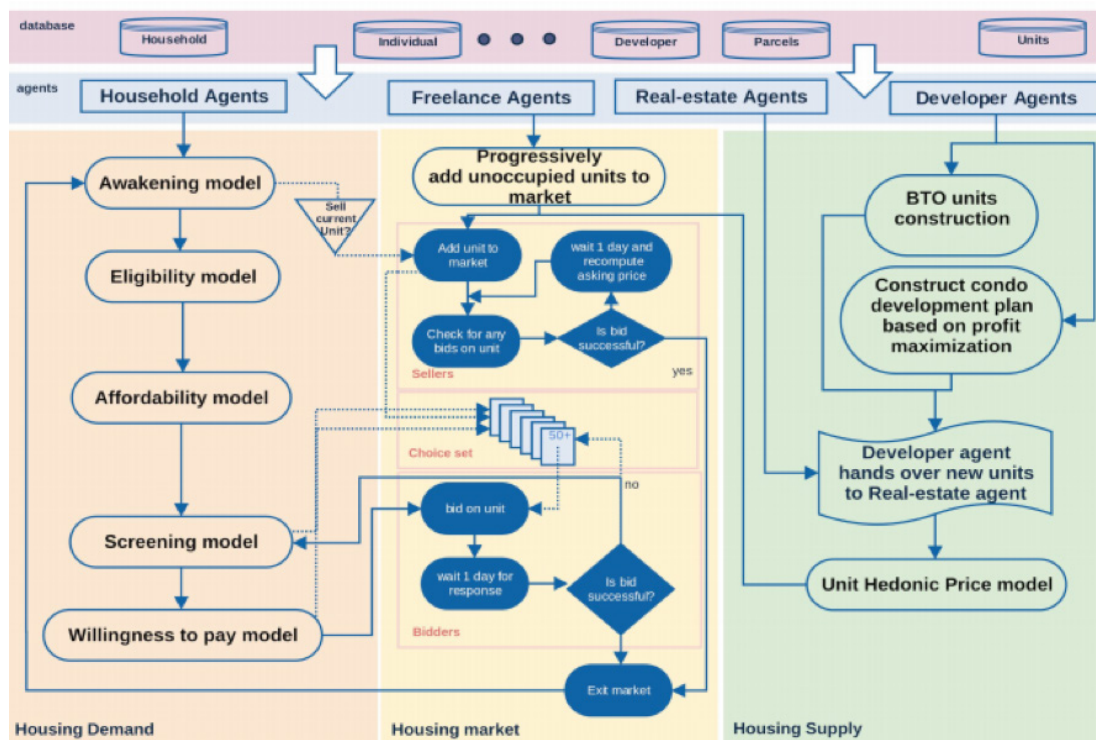
Figure 3: Flowchart of the dynamic bidding process approach in housing market

5.1   Residential real estate market demand dynamics

The awakening model determines the likelihood, on any particular day, that a household not currently active in the housing market will begin searching for alternative housing. Moving rates [6] vary by household demographics and tenure. Households headed by younger adults tend to move more frequently than senior-headed households. Renters move more frequently than owners and a small fraction of moves involve tenure changes (from own to rent or rent to own). The awakening model was calibrated using Singapore census data plus a "recent mover" survey of 6000 households, including 1000 who had moved at least once in the prior three years.[7]

Table 4 reports the moving rates and tenure transition probabilities currently used in our simulations. We distinguish households by tenure and age-of-head and assume that tenure transition probabilities are independent of moving rates. Hence, homeowners in the youngest household group have a 20% chance of moving within a year and a 94% chance that they will remain a homeowner. However, renters in the youngest household group have a 40% chance of moving per year and a 20% chance that they will continue to rent their housing.

---

[6]Awakening rates are higher than observed moving rates since they represent the start of a search for alternative housing and not all searches are successful. For example, for demographic group with a moving rate of 10% per year and half of all searches being successful, the awakening rate would be 20% per year.

[7]The recent mover survey was retrospective (i.e., asking about prior moves) and resulted in a sample biased towards younger and richer households; analysis is complicated by the fact that moves may involve changes in family structure (Shaw, 2018). The results are not yet fully incorporated into our current simulator which uses estimated moving and tenure transition rates that account for tenure changes and for sampling and retrospective biases but do not yet address family formation.

Table 4: Moving and tenure transition rates

Household demographics	Current tenure	Moving rate	Tenure transition	Transition rate (%)
Youngest (Head <35)	own	20	own-own	94
Youngest (Head <35)	own	20	own-rent	6
Young (Head 35-49)	own	10	own-own	95
Young (Head 35-49)	own	10	own-rent	5
Middle (Head 50-64)	own	6	own-own	93
Middle (Head 50-64)	own	6	own-rent	7
Senior (Head >=65)	own	4	own-own	89
Senior (Head >=65)	own	4	own-rent	11
Youngest (Head <35)	rent	40	rent-own	80
Youngest (Head <35)	rent	40	rent-rent	20
Young (Head 35-49)	rent	20	rent-own	93
Young (Head 35-49)	rent	20	rent-rent	7
Middle (Head 50-64)	rent	11	rent-own	87
Middle (Head 50-64)	rent	11	rent-rent	13
Senior (Head >=65)	rent	8	rent-own	71
Senior (Head >=65)	rent	8	rent-rent	29

The current implementation in Singapore focuses on home ownership rather than rentals since over 80% of Singapore citizens and permanent residents live in owner-occupied housing. Although the awakening model accounts for tenure and tenure transition, we only simulate the search and bidding behavior of awakened households (renters or owners) who seek to purchase housing. Our bidding approach should work for rental markets as well, and we intend to develop that model in the future.

On any given day, awakened households and any households currently in limbo (not assigned to a dwelling unit) constitute the market's potential buyers. The total number of households awakened each day is a user-defined parameter (which we estimate based on annual sales and rough estimates of search time and bidding success rates). The awakening likelihoods are multiplied by the overall daily awakening rate to determine the probability that an un-wakened household will be awakened on any particular day. Once awakened, a household will continue searching until it bids successfully on a housing unit, or gives up after a number of unsuccessful days (currently set to 90 days). Certain households are prevented from being awakened—e.g., all successful bids take at least 30 days to close and advance purchase of unfinished units might delay occupancy for up to four years. Households with such pending sales cannot be (re)awakened. Similarly, unsuccessful bidders have a cooling off period before returning to the pool of potentially awakened households.

Once a household is awakened, the simulator determines the maximum price that the household can afford to pay (based on monthly income, number and age-of-workers, and the current value of one's housing unit if it is owned), and whether the household is eligible for government subsidies if it purchases certain types of public housing units. Only affordable [8] and eligible housing units can be chosen from a household's "choice set."

---

[8] To be allowed in a household's choice set, the asking price of a housing unit must be no more than a few percent above the household's affordability threshold, and any bid cannot exceed the affordability limit.

On any given day, a household's choice set is formed by drawing 50+ units probabilistically from all available for-sale housing units. The likelihood that a housing unit is included in the choice set of an awakened household is proportional to the screening probabilities that estimate the odds of that household choosing to live in the planning district and housing type that match the candidate housing unit. These odds ratios come from a multinomial logit model estimated on data from a housing survey. This approach attempts to mimic the typical search process whereby households tend to narrow their search to particular neighborhoods and housing types. The approach also guards against including too many unlikely candidates in the choice sets. In addition, up to 10 new public housing units (offered by the Housing and Development Board [HDB]) can be added to a household's choice set (beyond the 50 private and HDB resale units) if such housing is available and the household is eligible. [9]

Once an awakened household's choice set is constructed, the household determines its "willingness to pay" for each possible housing unit and the optimal bid that would maximize the expected utility of the unit (including the likelihood that the offer is accepted by the seller). The approach follows Rocco (2014) who builds upon Lerman and Kern (1983). Households are heterogeneous in their valuation of housing bundle characteristics, and the market prices represent some amalgamation of the preferences of those types of households that tend to win the bidding for particular types of locations and housing units. Among candidate units, a household bids for the unit maximizing the surplus utility relative to its currently occupied unit. If no unit in the choice set gives positive surplus, the buyer does not bid. If the seller rejects the buyer's bid, the buyer will consider forgoing some utility surplus if re-bidding for the same unit on a subsequent day looks promising. The number of days buyers remain actively searching for housing is capped at a few months out of every year and households may forgo surplus or choose to rent if, after several weeks of searching, they are unsuccessful in bidding for suitable housing.

5.2   Real estate supply dynamics

Real estate supply currently has two basic agent types, developers and sellers. Developers' decisions include when, where, what type, and how much built space to build. In simulating these, LT uses models to predict future revenues of development projects, future construction cost, and a development-template-choice model to identify the optimal development template for a parcel including the option to hold the land undeveloped. The latter aspect is represented through real options theory, widely used in the real estate field to explain developers' behavior under market uncertainty. Sources of market uncertainty include future economic growth, population growth, social and cultural evolution, and real estate market dynamics. According to real options theory, vacant land can be considered a "call" option, which gives its owner the right, without obligation, to build a rent-producing structure upon paying construction costs. The option value of the land originates from the flexibility to expand investment, abandon development plans, defer investment for a period of time, etc. in response to anticipated market conditions. The developer will only develop the land when the expected investment return is higher than a threshold level. The greater the market uncertainty, the greater the value of the options to invest, and the higher the threshold return level for development.

The LT simulator models developers' investment decision as a development template choice for the parcel to maximize profit. A development template is characterized by a unique combination of property types (condominium, apartment, etc.), unit type (defined by unit sizes) mixture, and density, which are exogenously determined. These characteristics specify a development scenario that can be deployed to parcels, depending on development constraints such as zoning regulations on the parcel. A development template is only chosen by the developer if the expected investment return is greater than that of

---

[9] The HDB tends to offer 12,000 to 25,000 new units each year to eligible citizens and permanent residents.

all other development options and above a threshold value. Diao, Ma, and Ferreira (2015) provide more details on the relevant models.

Housing unit sellers set asking prices close to their expected market prices and, on any given day,[10] accept the highest bid exceeding their reservation price. The expected market price is estimated with a hedonic model (which is also used as part of the developer's model for new development decisions), detailed in the following section. Sellers consider reducing the asking price of unsold units for a specified time (generally weeks), until reaching the reservation price, and consider temporarily withdrawing units from the market if they do not sell within several months. An accepted bid results in the household being assigned the "pending unit"; the household has a pre-defined grace period for moving, which could be as large as four years (e.g., for advance purchase of built-to-order HDB new units). A number of parameters are exogenously set for factors such as: asking price adjustments for units that remain unsold for a week, bid adjustments when a bid is not accepted but the unit remains unsold, the period successful bidders wait before moving, and various values for buyers' and sellers' "time-on-market" and "time-off-market."

5.3   Real options-based development model

Real estate price plays a key role in developers' investment decision. LT currently uses a hedonic-price-model-based approach to predict the expected future revenue of a proposed development project based upon expected date of completed construction:

$$P^{gj(t)} = \varphi^j + \sum_{m=1}^{M} \phi^{mj} H^{mgj(t)} + \sum_{n=1}^{N} \chi^{nj} L^{ngj(t)} + \sum_{t=1}^{T} \tau^{j(t)} D^{g(t)} + \varepsilon^{gj(t)} \qquad (4)$$

where: $P^{gj(1)}$ is the log of transaction price of a housing unit g of unit type j in time t; H is a set of M property-specific attributes; L is a set of N locational amenities; D is a set of transaction time dummy variables reflecting housing market temporal dynamics; ε is an error term; and, ϕ, χ, and τ are vectors of parameter estimates.

The expected price of a housing unit upon the completion of construction consists of two parts. The expected value of housing attributes and locational amenities can be computed using the coefficients ϕ and χ through estimating Equation 4. The general housing market situation at the time of completion can be predicted based on historical market dynamics and macroeconomic factors:

$$\hat{\tau}^{j(t)} = \alpha_0^{j(t)} + \sum_{l=1}^{L} \beta_l^{j(t)} \hat{\tau}^{j(t-T_0-l+1)} + \sum_{k=1}^{K} r_k^{j(t)} Macro_k^{(t)} + e^{j(t)} \qquad (5)$$

where: $\hat{\tau}^{j(t-T_0-l+1)}$ is a set of L lagged terms of $\hat{\tau}^{j(t)}$ that starts from time $t-T_0-L+1$ to time $t-T_{0j}$; T_0 is construction time; L is determined based on the dataset used to calibrate the model; $Macro_k^{(t)}$ is a set of K macroeconomic factors at time t, such as GDP growth rate, interest rate, etc.; and, $\alpha_0^{j(t)}$, $\beta_l^{j(t)}$, and $r_k^{j(t)}$ are coefficients to be estimated.

One challenge in simulating real estate prices is the long-term pricing formation and evolution process. The hedonic price model specification (Equation 4) is estimated from historical housing transactions, assuming that household willingness-to-pay for housing attributes and amenities remains constant over time. Ignoring this (possibly strong assumption), prices still evolve based on changes in locational amenities (N in Equation 4) and housing types (H in Equation 4). Therefore, the simulation still needs a mechanism to formulate the overall housing market dynamics, so that the price prediction can be extended into the future. We generate the housing market dynamics as captured by $\hat{\tau}^{j(t)}$ in

---

[10]At "midnight" each seller reviews all bids and accepts the highest bidder for each unit as long as the bid exceeds a reservation price.

Equation 5 in future years based on simulated housing transactions as described below.

The LT simulator includes a dynamic housing bidding model to simulate the bargaining process between individual buyers and sellers as described in Section 5.2, which entails a bargaining dynamic with endogenously changing asking prices and bids and subsequent housing transactions. Assuming there are G simulated transactions of unit type j in time t of the simulation with known housing and location attributes, $\hat{\tau}^{j(t)}$ can be computed as:

$$\hat{\tau}^{j(t)} = \frac{1}{G}\sum_{g=1}^{G}(P^{gj(t)} - \hat{\varphi}^{j} - \sum_{m=1}^{M}\hat{\phi}^{mj} H^{mgj(t)} - \sum_{n=1}^{N}\hat{\chi}^{nj}L^{ngj(t)}) \qquad (6)$$

where: $P^{gj(t)}$ is the log of the simulated transaction price of property g of unit type j at time t; $H^{mgj(t)}$ and $L^{ngj(t)}$ are vectors of property and locational attributes, respectively; and $\hat{\varphi}^{j}$, $\hat{\phi}^{mj}$ and $\hat{\chi}^{nj}$ are estimated coefficients from the hedonic price model (Equation 4). In this way, the market dynamics as captured by $\hat{\tau}^{j(t)}$ can be updated continuously based on simulated transactions in the market over the entire simulation period.

5.4   Operationalizing activity-based accessibility (ABA) in the real estate market models

As indicated in Table 3, ABA measures figure into several of the LT behavioral models. Currently the models use ABA values calculated for the TAZ. Calculating ABA measures for these purposes requires considerable computational effort. First, SimMobility MT is iterated until travel times stabilize, so that each simulated individual's pre-day activity plan uses network performance characteristics consistent with the pre-day plan. This "quasi-equilibrium" reasonably represents an individual's accessibility expectations when evaluating alternative housing, work, and vehicle ownership options. For example, consider the housing bidding model. Suppose an individual considers relocating their residence from one TAZ to another while retaining the same workplace and vehicle access. If this person is the only one who moves, then the system performance of the network will be little changed and we need only evaluate the accessibility expected from the (possibly changed) pre-day activity plan if this individual were to move to the new residential location. In this way, we can generate an ABA estimate for every individual if they were to move to any of the 1169 TAZs while retaining the same workplace and vehicle access. In the current model implementation, this function call to the mid-term model averages around 0.2 seconds to execute and we pre-compute and store these giant tables, for every simulated individual and TAZ, for: fixed workplace-and-vehicles with varying residence; fixed residence-and-vehicles with varying workplace, and fixed workplace-and-residence with varying vehicle ownership.

Currently, we use a simplified method for generating household accessibility measures from the individual ABAs. Since we have not yet developed satisfactory (time or cost-based) scaling factors (see Dong, Ben-Akiva, Bowman, & Walker, 2006) for the logsum accessibility measures, we approximate a household's accessibility by using that of its highest income worker (or, if the household has no worker, the ABA of its member with the highest logsum). The location attributes in the hedonic model (Equation 4) estimation includes accessibility as the unit's TAZ average ABA measure of all households, assuming they chose to reside in that TAZ.[11] In principle, the willingness-to-pay model would use the household-specific accessibility measure rather than a TAZ average, which would help differentiate a household's valuation of a housing unit from the market price. However, current computational constraints in the simulation and data limitations for estimating willingness to pay made this impractical and we thus far use the same TAZ-level averages used in the hedonic model of market price.

---

[11]This resulted in a better fit then if we weighted households by the likelihood that they resided in the TAZ.

6 Supply side model estimation results

This section presents estimation results of the hedonic price model (Equation 4) and housing market forecast model (Equation 5) for Singapore. The housing transaction data used in estimation come from the REALIS database, a real estate information system managed by Singapore's Urban Redevelopment Authority (URA). The dataset contains detailed records of private housing transactions in Singapore from 1995 to 2017, including transaction price and date, street address, postal code, and various property attributes (e.g., floor area, floor level, age, property type, lease type, sale type). Locational (dis) amenities include distance to the nearest MRT station, distance to the nearest shopping mall, distance to the top 30 primary schools, distance to the nearest bus stop, and distance to the nearest expressway. We classify private housing units into 14 types, according to property type and floor area, and estimate separate models for each unit type.

6.1 Results

Table 5 shows the estimation results of the hedonic price model by unit type for condominium transactions. The estimated coefficients show that the market's willingness-to-pay[12] for housing and location attributes varies across unit types, but generally fit our expectations. The log-sum ABA has a positive and significant coefficient in all models, providing some evidence to support the effectiveness of this measure in capturing the heterogeneity in accessibility across locations. The models show reasonably good explanatory power.

This approach assumes that developers know both the housing and locational attributes of housing units in a new development project and the market's willingness-to-pay for these attributes. For the developer, therefore, the value of housing unit and its locational attributes is deterministic. The developers need only forecast the general housing market situation ($\hat{\tau}^{j(t)}$ in Equation 5) to predict future revenue of a proposed development project at time of completion. Assuming that the construction time of a development project is 1 year, we regress $\hat{\tau}^{j(t)}$ (estimated from the hedonic price model) on its lagged terms and macroeconomic factors. Table 6 presents the estimation results by unit type for condominium transactions. The results suggest that the market condition in a quarter is a good predictor of the market condition one year after. In addition, the GDP growth rate and real interest rate tend to have positive and significant effect on expected housing market condition.

---

[12]The hedonic model represents our estimate of market price for a housing bundle. The coefficients of the various terms (in Equation 4) can be thought of as reflecting a market average willingness-to-pay for housing characteristics and locational amenities across those households who win the bidding for housing in that neighborhood. Particular households will value a housing bundle differently. The term "market willingness-to-pay" refers to the valuation of locational amenities (including accessibility). Note that, when particular households bid for particular housing units in our bidding model, they use a willingness-to-pay estimate that is specific to their household characteristics and workplaces. Hence, their willingness-to-pay equation has coefficients that may differ from those in the hedonic model of market price.

Table 5: Estimation results for hedonic price models (condominium)

Unit type	Condo (<60m2)	Condo (60-79m2)	Condo (80-99 m2)	Condo (100-134m2)	Condo (>=135m2)
log(area)	0.932***	0.864***	0.916***	1.105***	1.056***
	(0.011)	(0.018)	(0.016)	(0.009)	(0.005)
Freehold	0.155***	0.155***	0.171***	0.146***	0.207***
	(0.005)	(0.006)	(0.004)	(0.003)	(0.005)
Age	-0.004***	-0.010***	-0.019***	-0.020***	-0.030***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Age^2	0.000***	0.000***	0.000***	0.000***	0.001***
	(0.000)	(0.000)	(0.000)	0.000	(0.000)
Age missing	0.057***	-0.089***	-0.071***	-0.079***	-0.171***
	(0.007)	(0.008)	(0.005)	(0.003)	(0.006)
Storey	0.001	-0.002***	0.003***	0.006***	0.008***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Storey^2	0.000***	0.000***	0.000***	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-19.557***	-22.528***	-29.622***	-27.044***	-35.864***
	(0.403)	(0.411)	(0.277)	(0.202)	(0.371)
Logsum accessibility	5.538	6.205***	7.512***	6.834***	8.578***
	(0.076)	(0.076)	(0.051)	(0.038)	(0.072)
Within 1km of a top primary school	0.029***	0.040***	0.009***	0.015***	-0.037***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)
Distance to major shopping mall (km)	-0.027***	-0.043***	-0.025***	-0.024***	-0.071***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Within 200m of a MRT station	-0.016***	-0.078***	-0.018***	-0.062***	-0.116***
	(0.006)	(0.005)	(0.004)	(0.003)	(0.008)
Within 200-400m of a MRT station	0.035***	0.042***	-0.006**	-0.018***	0.013***
	(0.004)	(0.004)	(0.003)	(0.002)	(0.005)
Within 200m of expressway	-0.086***	-0.039***	-0.037***	0.016***	-0.027***
	(0.004)	(0.004)	(0.003)	(0.002)	(0.004)
Within 200-400m of a bus stop	0.065***	0.017***	0.046***	0.041***	0.100***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)
Beyond 400m of a bus stop	0.005	0.138***	0.116***	0.238***	0.610***
	(0.018)	(0.012)	(0.008)	(0.006)	(0.006)
Transaction quarter	Yes	Yes	Yes	Yes	Yes
Observations	11,176	13,911	35,011	78,464	55,740
R^2	0.767	0.66	0.72	0.679	0.742
Adjusted R^2	0.765	0.658	0.72	0.678	0.742

Notes: *** $p<0.01$; ** $p<0.05$ * $p<0.1$

Table 6: Estimation results for housing market prediction models (condominium)

Unit type	Condo (<60m2)	Condo (60-79m2)	Condo (80-99 m2)	Condo (100-134 m2)	Condo (>=135m2)
$\hat{\tau}^{j(t-4)}$	0.724***	0.800***	1.395***	1.489***	1.739***
	(0.195)	(0.224)	(0.324)	(0.335)	(0.321)
$\hat{\tau}^{j(t-5)}$	0.131	0.253	-0.363	-0.563	-0.767
	(0.202)	(0.252)	(0.466)	(0.542)	(0.473)
$\hat{\tau}^{j(t-6)}$	0.268	-0.081	-0.511	-0.500	-0.663
	(0.209)	(0.253)	(0.458)	(0.538)	(0.474)
$\hat{\tau}^{j(t-7)}$	-0.106	0.198	0.504	0.634*	0.815**
	(0.195)	(0.213)	(0.316)	(0.330)	(0.305)
Quarterly GDP growth rate	0.005*	0.005**	0.005**	0.004**	0.008***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Real interest rate	0.037	0.080**	0.028	0.038	0.096***
	(0.040)	(0.031)	(0.026)	(0.024)	(0.031)
Constant	-0.036	-0.086*	-0.023	-0.029	-0.130**
	(0.083)	(0.045)	(0.040)	(0.038)	(0.056)
Observations	54	54	54	54	54
R²	0.669	0.760	0.789	0.812	0.759
Adjusted R²	0.627	0.729	0.762	0.788	0.728

Notes:*** p<0.01;** p<0.05* p<0.1

7   Simulation example: New public housing offering

This section demonstrates LT's capabilities—specifically, the housing market model—to simulate the spatial and economic impacts of new public housing offerings. During 2012, the HDB launched 26,252 new housing units as part of its Built to Order (BTO) program (Housing and Development Board, 2013). For more than a decade, the HDB has offered 10,000-30,000 such units each year depending on anticipated population growth and "market calming" policies. Most such units are part of "new estate" public housing developments in the expanding suburbs. The 2012 offerings were bundled into six offerings of around 4,000 units each within 20-50 story towers typically containing hundreds of units. HDB sets the prices. Eligible households have one month to bid for specific units and a lottery determines the winners. If more than 75% of the units in a building have bids, then construction commences—hence the term "built to order."

We simulate the year 2012 with and without the BTO offerings. Table 7 summarizes the initial population and spatial entities used. The estimated accessibility measures were calculated based on the simulated travel for the entire Singapore population, but the households actively searching in the housing market excluded non-permanent residents (and renters).

Table 7: 2012 synthetic population and spatial detail

Individuals*	4,040,920
Households*	1,149,325
Establishments	169,288
Housing units	1,387,774
Buildings	171,654
Land parcels	184,322
Postcodes	116,468
Traffic Analysis Zones	1,169
Planning Subzones	312
Planning Areas	55

* Housing market simulation excludes 1.3 million construction workers, work permit holders and other foreign households.

Table 8: Configuration parameters – bidding parameters

Bidding parameters	Value
Initial vacancy rate (owner-occupied)	3.0%
Daily household awakenings	400 days
Max HH time actively searching	90 days
Min HH time off-market	180 days
Max time on market if unsold unit	210 days
Time off market if unit not sold	210 days
Move in time after winning bid	30 days
Alert percent if BTO-eligible	2.5%

The moving rates and tenure transition rates were estimated based on census data and a "recent mover" survey as described in Section 5.1 (see Table 4). We assume that the own-rent transition probabilities are independent of the decision to move. The rates are preliminary since the analysis of the recent mover survey has only recently been completed (Shaw, 2018). For the simulation presented, we made assumptions on the configuration parameters in Table 6 in light of available data such as housing transactions and our understanding of local housing mobility. The overall ownership vacancy rate is derived from the integration of households and housing units in our synthetic population. The 400 households awakened each day are sampled from households (who have not moved too recently) using moving rates and ownership transition likelihoods from Table 4. We are in the process of tuning these parameters plus the time-to-closing and time on/off market parameters, and several bidding model parameters so that we can match various literature estimates and market totals that are available for 2012 regarding total sales, bidding success rate, average search time, and the like.

For the comparison, we ran 10 simulations of one year without BTO sales and 10 simulations with the BTO units as actually offered during six separate months of 2012. For the BTO runs, we awakened an additional 2.5% of the BTO-eligible households during each day that BTO units were available for bidding. Households can 'subscribe' to a messaging service that alerts them at the start of a day if market conditions have changed. Alerting 2.5% households per day yields an overall number of awakened and BTO-eligible households that approximates the observed number of BTO bidders reported in HDB statistics.

For awakened households in both the baseline and BTO runs, the same affordability, eligibility, and filtering steps are used to construct the household's choice set. The bidding and selling agents' behaviors follow the processes described in Sections 5.1 and 5.2. Figure 4 (top) shows the counts of BTO housing units that were offered in each of Singapore's 57 planning districts during 2012. The bulk of the 25,000 units were offered in non-mature suburban estates in the northeast and northwest. The 10 runs with BTO sales averaged 6 thousand fewer non-BTO sales than the 10 baseline runs without any BTO sales, suggesting that 24% of the BTO sales were substitute sales by households who might otherwise have purchased a private or HDB resale unit. The average sales price of the non-BTO units did drop slightly when the BTO units were offered, but only by 1%. This slight drop is not unexpected since the 6,000 substitute sales are less than 10% of the annual sales and many additional households were "awakened" only because of the BTO offering and would not otherwise have been actively searching for new housing. Figure 4 (bottom) shows the percentage reduction in sales among the planning districts for the 10 non-BTO simulations compared with the 10 BTO simulations. Only planning districts that had a statistically significant difference in sales are highlighted. Note that most affected districts are those close to the districts where most of the BTO offerings are located.

While preliminary, the results illustrate the types of price, substitution, and spatial effects that motivate our interest in developing a daily bidding model for real estate markets that has rich spatial detail and is tightly integrated with an activity-based transportation model.[13]

---

[13]As indicated in Table 5, ABA measures play a significant role in predicting the market prices on which sellers base their asking prices. Likewise, although the results are not shown here, ABA is also significant and positive in buyers' willingness-to-pay.
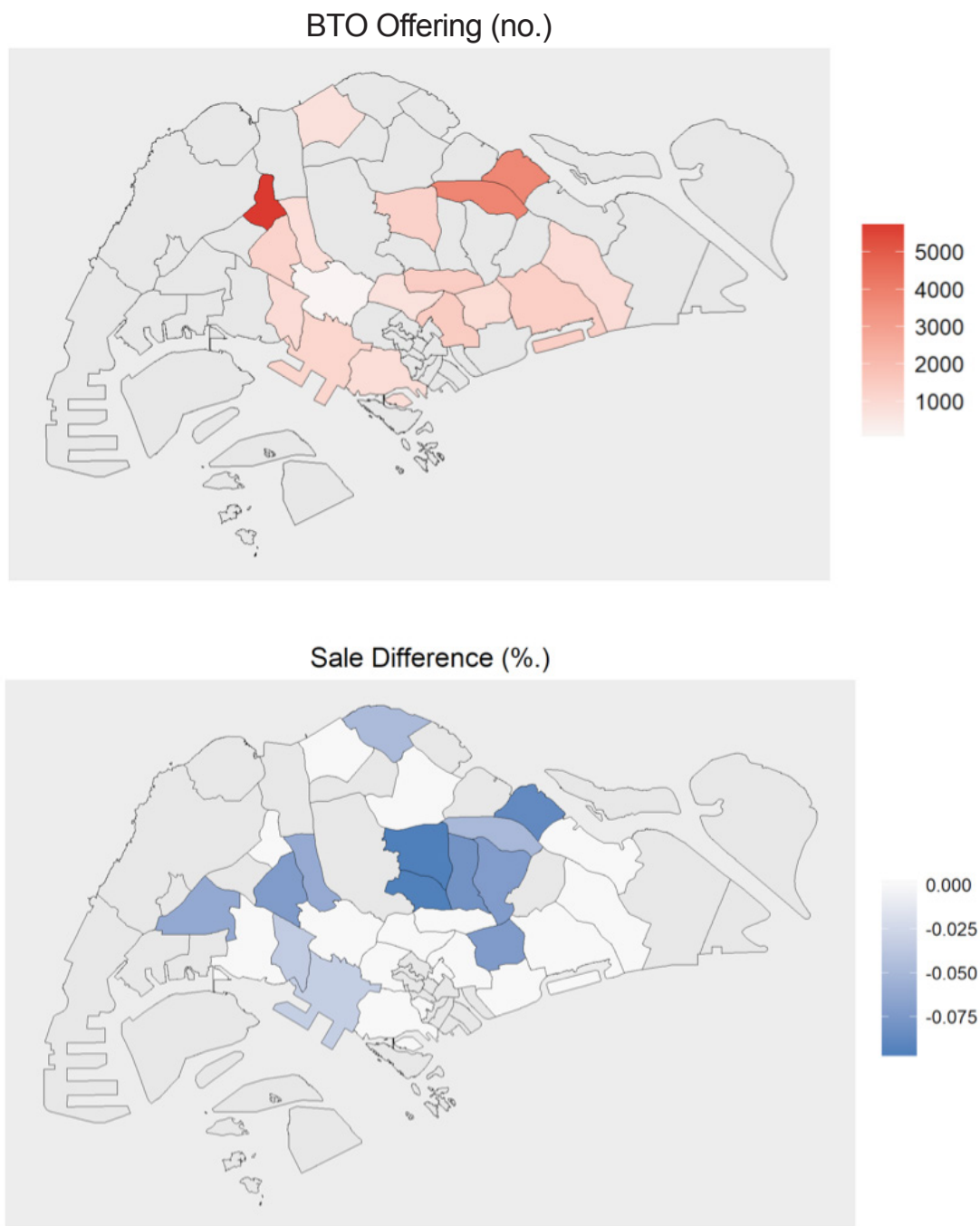
Figure 4: BTO units offered during 2012 (top) and percent difference in BTO and non-BTO sales (bottom) (by Singapore Planning District)

8   Conclusions and future work

In the past 20 years, integrated land-use and transport models have been moving from being aggregate, equilibrium-based to disaggregate, dynamic, and microsimulation-based. Integrated microsimulation has numerous advantages, in theory, such as providing the possibilities to: better capture the behavioral richness reflected in state-of-the-art activity based models of travel demand; model specific types of rules,

regulations and/or prices targeted at specific agent types in specific situations; and, capture individual agent behavior and behavioral interactions (among agents). This paper presents one such integrated microsimulation tool, SimMobility, focusing on its "long-term" (LT) simulator. Unlike most other attempts at integrating microsimulation-based land-use and transport models, SimMobility explicitly uses activity-based accessibility (ABA) measures to reflect how longer-term choices, such as where to live, relate to daily behaviors, such as where and when to shop and how to get there. Utility-based ABA provides a theoretically, behaviorally and micro-economically consistent means of integrating land-use and transportation models.

The LT simulator models a range of individual, household, and firm decisions that tend to be made over the period of months and years, such as the choice of job and school location, vehicle ownership, residential location, and real estate development. This paper details LT's approach to modeling the housing market, which includes: (a) "awakening" households to begin searching for new housing; (b) handling eligibility and affordability constraints; (c) screening the available supply to identify a likely choice set; and, (d) modeling real-estate developers' decisions regarding when, where, what type, and how much built space to construct, accounting for market cycles and uncertainty. A simple example showed how the LT simulator can be used to examine the spatial and economic impacts of urban land use and transportation policies and options. Two sets of 10 simulations were run for the year 2012 in Singapore, with and without consideration of new public housing supply. The simulations allowed for a comparison of the estimated substitution, price, and spatial impacts of the public housing program.

While still a work-in-progress, SimMobility LT demonstrates the feasibility of simulating daily transactions in the housing market in a manner that integrates ABA measures in a theoretically consistent, yet computationally feasible way. The relevant LT submodels make the daily bidding simulations feasible, with: (a) model specifications that can be readily enhanced with better market data and (b) behavioral parameters that would potentially allow testing of, for example the effects of various marketing programs and incentives to promote car-lite communities or subscription to car sharing services.

Despite its promise, SimMobility's complexity and the tight interdependencies between its LT and MT modules pose a range of challenges. Singapore is relatively "data rich," with good base maps; bus, taxi, and public transit monitoring; travel surveys; real estate transaction databases, etc. Still, estimating the behavioral models, much less calibrating and validating the simulations, requires a time-consuming, delay-prone process of data acquisition, cleaning, and spatial and temporal integration. Even then, the "right" data remain elusive for some of the models (e.g., MT's activity-based models, LT's bidding models). Model enhancements in one module could imply major changes in another. For example, Jiang (2015) shows that latent class choice models for household location in Singapore improve on the standard (one-class) case, reflecting how sensitivities to accessibility vary across class type. Proper integration of such an approach with MT, however, would require that MT add the latent class distinctions to (at least) its pre-day activity plan model specification for individuals. This would then imply major time to re-estimate the model and recalibrate its interactions with the other demand and supply components of the MT and ST simulators. Calculating the ABAs in the MT for LT model estimation and, then, simulation also consumes large amounts of computational resources.

Nonetheless, a range of factors suggests that integrated, activity-based microsimulation models, like SimMobility, will have a valuable potential role to play as we head into the future. Ongoing improvements in communication and logistics have already ushered in an era of sharing, whether it be vehicles, rides, and/or workspaces. E-commerce is changing shopping behavior and trip making and urban logistics. Driverless vehicles and on-demand mobility services will change demand for parking and the demand and supply of public transportation and, quite possibly, individuals' value of travel time. The impacts on housing and firm location choices, job locations, individual activity patterns, and the like are uncertain and yet important to know. Theoretically consistent, integrated and behaviorally explicit

microsimulation tools provide a good framework for examining these topics. That said, integrated microsimulation tools are not crystal balls; their capability to "predict" the future is only as good as their underlying model specifications and assumptions, exogenous inputs (e.g., economic conditions), data, behavioral representation (e.g., parameters and stability over time), among other factors. More sophisticated microsimulation tools do not, on their own, overcome the challenges inherent; indeed, they may compound them. Used creatively, however, they can even be used to better understand the implications of their underlying uncertainties (e.g., Goodings Swartz, & Zegras, 2013). Increasing computational power may make their use towards such ends ever more feasible.

Enhancing, calibrating, and validating our current LT simulator using Singapore data from 2008 through 2016 remain ongoing tasks. At the same time, various portions of SimMobility are already being tested in other contexts (e.g., Greater Boston). Relevant future research needs include:

- Improving the current approach to handling year-to-year changes in the total number of people and firm establishments in a way that manages the "history" needed to reflect relevant temporal constraints and lifecycle events that impact household behavior yet avoids constructing a full-blown demographic model (e.g., of births, deaths, family formation);
- Better reflecting household and individual lifecycles and lifestyles and their implications for the housing market (e.g., relocations, tenure transition) and their inter-relationships with vehicle ownership, etc.;
- Extending the housing market model to include a bidding model for rentals;
- Capturing the investment dimensions of homeownership rather than just modeling occupancy;
- Improving the willingness-to-pay model to differentiate the value of particular housing bundles to households with different characteristics;
- Moving to a vehicle transaction model that reflects cost of ownership, subscription services (e.g., to vehicle sharing services), and new modes such as personal mobility devices;
- Expanding the real estate market sub-module to handle non-residential development and to model firm/establishment relocations;
- Exploring more efficient algorithms to improve the computational performance of the LT simulator and further facilitate the data exchange between LT and MT;
- Testing if the theoretical advantage of using the activity-based measures is "worth it" in practice (i.e., if it improves model performance relative to more traditional accessibility measures);
- Understanding the implications of stochastic variation, model misspecification, imperfect input data, and the innate randomness in the models to improve strategies for maximizing what we learn from model runs and minimizing misinterpretation of results; and,
- Assessing whether microsimulation, and the costs and efforts it implies, is really better than more traditional approaches.

In the almost half century since Lee's (1973) requiem for large scale models, such models are alive and thriving, if not necessarily in a new "golden age." SimMobility exemplifies the movement towards integrated microsimulation approaches to urban modeling. While much has been achieved, much work remains in ensuring such tools can really help contribute to improved metropolitan decision-making in the 21st century.

Acknowledgements

References

Adnan, M., Pereira, F. C., Lima Azevedo, C. M., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C., & Ben-Akiva, M. E. (2016). SimMobility: A multi-scale integrated agent-based simulation platform. Paper presented at the Transportation Research Board 95th Annual Meeting, Washington, DC.

Ben-Akiva, M., & Bowman, J. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies, 35(7)*, 1131–1153.

Ben-Akiva, M., Bowman, J., & Gopinath, D. (1996). Travel demand model system for the information era. *Transportation, 23*(3), 241–266.

Ben-Akiva, M., & Lerman, S. (1979). Disaggregate travel and mobility-choice models and measures of accessibility. In D. Hensher & P. Stopher (eds.) *Behavioral travel modelling* (pp. 654–679). Proceedings of the 3rd International Conference on Behavioral Travel Modeling, Tanunda, Australia, 1977.

Bhat, C. R., & Singh, S. K. (2000). A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A, 34*(1), 1–22.

Buliung, R., & Kanaroglou, P. (2007). Activity-travel behavior research: Conceptual issues, state of the art, and emerging perspectives on behavioral analysis and simulation modelling. *Transport Reviews, 27*(2), 151–187.

Castiglione, J., Bradley, M., & Gliebe, J. (2015). Activity-based travel demand models: A primer. TRB, AHRP 2 Report (S2-C46-RR-1). Retrieved from https://dx.doi.org/10.17226/22357

Charypar, D., & Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation, 32*(4), 369–397.

Diao, M., Ma, X., & Ferreira, J. (2015). A real-option based dynamic model to simulate real estate developer behavior. *Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management,* Cambridge, MA.

Dong, X., Ben-Akiva, M. E., Bowman, J. L., &Walker, J. L. (2006). Moving from trip-based to activity-based measures of accessibility. *Transportation Research, Part A, 40*(2), 163–180.

Donnelly, R., Upton, W., & Knudson, B. (2018). Oregon's transportation and land-use model integration program: A retrospective. *Journal of Transport and Land Use, 11*(1), 19–30.

Geurs, K. T., & Van Wee., B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography, 12*(2), 127–140.

Goodings Swartz, P., & Zegras., P. C. (2013). Strategically robust urban planning? A demonstration of concept. *Environment and Planning B, 40*(5), 829–845.

Housing and Development Board. (2013). HDB key statistics 2012. Retrieved from http://www10.hdb.gov.sg/eBook/AR2012/keystatistics.html

Jiang, S. (2015). *Deciphering human activities in complex urban systems: Mining big data for sustainable urban future.* (Doctor dissertation.) Cambridge, MA: Department of Urban Studies and Planning, Massachusetts Institute of Technology. Retrieved from http://hdl.handle.net/1721.1/101369

Kii, M., Nakanishi, H., Nakamura, K., & Doi, K. (2016). Transportation and spatial development: An overview and a future direction. *Transport Policy, 49*, 148–158.

Lee, D. B. (1973). Requiem for large-scale models. *Journal of the American Institute of Planners, 39*(3), 163–178.

Lerman, S. R., & Kern, C. R. (1983). Hedonic theory, bid rents, and willingness to pay: Some extensions of Ellickson's results. *Journal of Urban Economy, 13*, 358–363.

Lu, Y., Adnan, M., Basak, K., Câmara Pereira, F., Carrion, C., Saber, V. H., Loganathan, H., & Ben-Akiva, M. (2015). Simmobility mid-term simulator: A state-of the-art integrated agent-based de-

mand and supply model. Paper presented at the Transportation Research Board 94th Annual Meeting, Washington, DC.

Martínez, F. J. (1995). Access: The transport-land use economic link. *Transportation Research Part B, 29*(6), 457–470.

Miller, E. J., Hunt, J. D., Abraham, J. E., & Salvini, P. A. (2004). Microsimulating urban systems. *Computers, Environment and Urban Systems, 28*(1), 9–44.

Nicolai, T., & Nagel, K. (2012). Coupling transport and land use: Investigating accessibility indicators for feedback from a travel to a land-use model. Paper presented at Latsis Symposium 2012–1st European Symposium on Quantitative Methods in Transportation Systems, Lausanne, Switzerland.

Pendyala, R., Konduri, K., Chiu, Y. C., Hickman, M., Noh, H., Waddell, P., & Gardner, B. (2012). Integrated land use-transport model system with dynamic time-dependent activity-travel microsimulation. *Transportation Research Record, 2303*, 19–27.

Rocco, V. (2014). *Microsimulación de transacciones en el Mercado Inmobiliario.* (Doctor dissertation.) Cambridge, MA: Department of Urban Studies and Planning, Massachusetts Institute of Technology.

Shaw, J. (2018). Modeling households' life cycle stages and residential relocation behavior. (Doctor dissertation.) Cambridge, MA: Department of Urban Studies and Planning, Massachusetts Institute of Technology.

Small, K. A., & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica, 49*, 105–130.

Waddell P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics, 3*(1), 43–67.

Waddell, P. (2000). Urbansim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association, 68*(3), 297–314.

Wagner, P., & Wegener, M. (2007). Urban land use, transport, and environment Models: Experiences with an integrated microscopic approach. *disP-the Planning Review, 43*(170), 45–56.

Wegener, M. (2004). Overview of land-use transport models. In D. A. Hensher & K. Button (eds.), *Transport geography and spatial systems* (pp. 127–146). Kidlington, UK: Pergamon/Elsevier Science.

Zhu, Y., & Ferreira, J. (2014). Synthetic population generation at disaggregated spatial scales for land use and transportation microsimulation. *Transportation Research Record, 2429*, 168–177.

Zhu, Y., & Ferreira, J. (2015). Data integration to create large-scale spatially detailed synthetic populations. In *Planning support systems and smart cities* (pp. 121–141). Cham, Switzerland: Springer International Publishing.

Zhuge, C., Shao, C., Gao, J., Dong, C., & Zhang, H. (2016). Agent-based joint model of residential location choice and real estate price for land use and transport model. *Computers, Environment and Urban Systems, 57*, 93–105.

Ziemke, D., Nagel, K., & Moeckel, R. (2016). Towards an agent-based, integrated land-use transport modeling system. *Procedia Computer Science, 83*, 958–963.